INTERNATIONAL JOURNAL OF ELECTRONICS AND COMPUTER APPLICATIONS

ORIGINAL ARTICLE

Emerging Advances in Multimodal Imaging and Fusion Techniques

Arati J Vyavahare[1]*, Diksha Anand Sawant[2]

1 Professor, ENTC, Modern College of Engineering, Pune-05, Maharashtra, India
2 Student, Modern College of Engineering, Pune-05, Maharashtra, India

Article access online

OPEN ACCESS

Received: 20.03.2025

Accepted: 10.05.2025

Published: 18.06.2025

Citation: Vyavahare AJ, Sawant DA. (2025). Emerging Advances in Multimodal Imaging and Fusion Techniques. International Journal of Electronics and Computer Applications. 2(1): 29-31. https://doi.org/10.70968/ijeaca.v2i1.D1019

***Corresponding author**.

arati.vyavahare@moderncoe.edu.in

Funding: None

Competing Interests: None

ISSN
Print:
Electronic: 3048-8257

Abstract

Multimodal image fusion is a rapidly evolving domain that combines complementary information from different imaging modalities into a unified representation, enhancing visual perception and decision-making in areas such as medical diagnostics, surveillance, and remote sensing. This paper surveys the taxonomy of multimodal fusion tasks, discusses classical and emerging deep learning-based methods including convolutional neural networks, autoencoders, generative adversarial networks, and transformer-based architectures and outlines key evaluation metrics. Through this comprehensive review, we highlight recent trends, performance benchmarks, and future directions for effective multimodal image fusion.

Keywords: Multimodal Imaging; Image Fusion; CNN; Autoencoders; GANs; Transformers; Medical Imaging; Feature Integration; Deep Learning; Evaluation Metrics

Introduction

Multimodal imaging involves capturing information using two or more imaging systems, each providing unique characteristics. For example, in medical imaging, MRI offers high spatial resolution, while PET reveals metabolic activity. The integration or fusion of such modalities leads to enhanced interpretation by combining structural and functional insights.

With the increasing availability and resolution of imaging technologies, image fusion has gained prominence. However, effective fusion poses challenges including modality misalignment, information redundancy, and loss of important details. Recent developments in machine learning and deep learning have significantly transformed how fusion tasks are approached.

This paper aims to provide a structured overview of this field by categorizing fusion tasks, comparing traditional and advanced deep learning techniques.

Taxonomy of Multimodal Image Fusion Tasks

Multimodal image fusion tasks can be categorized based on modality combinations, fusion strategy levels, and application domains. In terms of modality combinations, the most common examples include medical imaging (CT-MRI, PET-CT, PET-MRI, WSI-Omics), remote sensing (Panchromatic-Multispectral, SAR-Optical), and surveillance (Infrared-Visible). Fusion strategies operate at different levels: pixel-level fusion combines raw data, although it's often sensitive to noise and misalignment; feature-level fusion integrates intermediate features from neural networks or statistical models, offering greater robustness and flexibility[1–4]; and decision-level fusion merges the final outputs or predictions from each modality to form a consensus. As for applications, multimodal fusion finds critical use in clinical diagnosis and treatment planning (Pathomic Fusion, Diff4MMLiTS), survival prediction frameworks (MGCT, 4D-ACFNet), tumor microenvironment characterization[5,6], and radiation therapy planning[7–9].

Classical Approaches for Multimodal Image Fusion

Before the advent of deep learning, multimodal image fusion relied heavily on statistical and transform-based techniques. Spatial domain techniques such as simple averaging, Principal Component Analysis (PCA), and Intensity-Hue-Saturation (IHS) transformations offered basic fusion strategies. Transform domain methods included Discrete Wavelet Transform (DWT), Laplacian Pyramid decomposition, and Non-subsampled Contourlet Transform (NSCT), which allowed for multi-scale fusion representations. Optimization-based methods, including sparse representation and total variation minimization, aimed to preserve salient features while minimizing artifacts. Despite their computational efficiency, these classical approaches often lacked adaptability and were challenged by complex multimodal datasets.

Deep Learning Paradigms in Multimodal Image Fusion

Deep learning has significantly enhanced multimodal fusion by enabling automatic feature extraction and learning complex relationships. Convolutional Neural Networks (CNNs) such as DenseFuse, FusionNet, and Diff4MMLiTS[10] learn hierarchical features that capture local and global structures across modalities. Autoencoders (AEs) help compress and reconstruct fused representations by encoding inputs from different modalities and decoding them jointly to retain modality-specific features. Generative Adversarial Networks (GANs), like FusionGAN and IR-D-GAN, synthesize fused images that are visually realistic and maintain key textural and structural components. More recently, transformer-based models such as MGCT[11], TransFuse, and 4D-ACFNet[12] utilize attention mechanisms to model long-range dependencies and enhance cross-modal interactions. Collectively, these deep learning paradigms outperform classical methods in visual realism, modality preservation, and task-specific accuracy.

Evaluation Metrics for Fusion Quality

Evaluating the performance of fusion techniques is essential to measure their effectiveness. Common statistical and information-theoretic metrics include Entropy (EN), which reflects the information richness of the fused image, and Mutual Information (MI), which quantifies the shared information between source and fused images. Structural metrics such as Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Fusion Quality Index (Qabf) are frequently used to assess perceptual and quantitative fidelity. In clinical and task-based applications, subjective evaluations by human experts and the performance in downstream tasks such as cancer detection or survival prediction are equally important. Recent studies[3,4,13] advocate for hybrid evaluation frameworks combining statistical, visual, and task-specific metrics for a comprehensive assessment.

Conclusion and Future Directions

Multimodal image fusion has evolved into a crucial technology across domains such as medical imaging, surveillance, and remote sensing, offering enriched visual interpretation and improved decision-making by integrating complementary data from different modalities. This paper has provided a structured review of the field starting from classical fusion techniques to the transformative impact of deep learning methods such as CNNs, autoencoders, GANs, and transformers. Each paradigm contributes uniquely to feature extraction, modality preservation, and visual fidelity in fusion outcomes. Furthermore, we discussed the taxonomy of fusion tasks, key application areas, and evaluation metrics essential for assessing fusion quality.

As the field progresses, future research will focus on real-time and adaptive fusion systems, self-supervised learning, diffusion models, graph-based architectures, and interpretability to foster clinical trust and broader deployment. Through this review, we aim to support researchers and practitioners in navigating current advancements and identifying promising directions for innovation in multimodal image fusion.

Table 1. Comparison of Multimodal Image Fusion Techniques

Technique	Type	Key Characteristics	Accuracy	Adaptability	Computational Cost	Suitability for Complex Data
Simple Averaging / PCA / IHS	Classical	Basic pixel-level methods, fast but sensitive to noise and misalignment	Low	Very Low	Very Low	Poor
DWT / Laplacian / NSCT	Classical	Multi-resolution transforms, preserve edges and structures	Moderate	Low	Low	Moderate
Sparse Representation / TV Minimization	Classical	Optimization-based fusion, balances detail and noise	High	Moderate	Moderate	Moderate
Convolutional Neural Networks (CNNs)	Deep Learning	Automatically learn features, handle modality differences well	High	High	High	High
Autoencoders (AEs)	Deep Learning	Compress and jointly reconstruct representations from different modalities	High	Moderate	Moderate	Moderate to High
Generative Adversarial Networks (GANs)	Deep Learning	Produce visually realistic fused images with texture and structure preservation	Very High	High	Very High	High
Transformer-based Models	Deep Learning	Utilize attention to model long-range and cross-modal dependencies	Very High	Very High	Very High	Very High
Diffusion-based Models	Deep Learning	Robust to noise and missing modalities, data-driven image synthesis	Very High	Very High	High	Very High
Graph Neural Networks (GNNs)	Deep Learning	Represent spatial and relational information between modalities	High	High	High	High

References

1) Lv W, Ashrafinia S, Ma J, Lu L, Rahmim A. Multi-Level Multi-Modality Fusion Radiomics: Application to PET and CT Imaging for Prognostication of Head and Neck Cancer. *IEEE Journal of Biomedical and Health Informatics*. 2020;24(8):2268–2277. Available from: https://doi.org/10.1109/jbhi.2019.2956354.
2) Zhang Y, et al. Joint Modeling of Spatial Dependencies Across Multiple Subjects in Multiplexed Tissue Imaging. *IEEE Transactions on Medical Imaging*. 2025. Available from: https://arxiv.org/abs/2504.02693.
3) Bhardwaj S, et al. Revolutionizing Tumor Detection and Classification in Multimodality Imaging: Deep Learning Methods and Applications. *Proc IEEE Int Conf Biomed Imaging (ISBI)*. 2024;p. 103–107.
4) Zhang Y, et al. Attention-Map Augmentation for Hypercomplex Breast Cancer Classification. *IEEE Trans Med Imaging*. 2025;44(9):1456–1467.
5) Chen M, Wang L, Khosravi A, Navab N. AI-powered Interpretable Imaging Phenotypes Noninvasively Characterize Tumor Microenvironment in Breast Cancer. *NPJ Breast Cancer*. 2023;9(55):1–11.
6) Liu K, et al. Current Status and Future Prospects of Molecular Imaging in Targeting the Tumor Immune Microenvironment. *J Nanobiotechnol.* 2024;21(3):1–15.
7) Alavi M, et al. Artificial Intelligence in Four-Dimensional Imaging for Motion Management in Radiation Therapy. *Artif Intell Rev.* 2024;68(4):1245–1267.
8) Alavi M, et al. AI Applications in 4D Imaging for Radiation Therapy: State of the Art. *Proc IEEE Eng Med Biol Soc (EMBC)*. 2024;p. 567–571.
9) Alavi M, et al. Artificial Intelligence-Assisted Multimodal Imaging for the Clinical Applications of Breast Cancer: A Bibliometric Analysis. *IEEE Trans Med Imaging*. 2025;44(8):1345–1356.
10) Chen S, et al. Diff4MMLiTS: Advanced Multimodal Liver Tumor Segmentation via Diffusion-Based Image Synthesis and Alignment. *IEEE Trans Biomed Eng*. 2024;72(6):789–800. Available from: https://arxiv.org/abs/2412.20418.
11) Gu Y, Huang H, Tong Q, Cao M, Ming W, Zhang R, et al. Multi-View Radiomics Feature Fusion Reveals Distinct Immuno-Oncological Characteristics and Clinical Prognoses in Hepatocellular Carcinoma. *Cancers*. 2023;15(8):2338. Available from: https://dx.doi.org/10.3390/cancers15082338.
12) Li H, Zhou X, Zhang W, Fang J. 4D-ACFNet: A 4D Attention Mechanism-Based Prognostic Framework for CRLM. In: Proc. ACM Int. Conf. Multimedia (MM'25). 2025;p. 456–465.
13) Bhardwaj S, et al. Revolutionizing Tumor Detection and Classification in Multimodality Imaging Based on Deep Learning Approaches: Methods, Applications, and Limitations. *J X-ray Sci Technol*. 2024;31(2):233–245.